Hanyu Wang

why0725@mail.ustc.edu.cn · (+86) 180-366-70725 · https://wwwhy725.github.io/

Education

University of Science and Technology of China, Anhui, China — Sep. 2020 – Present

Undergraduate student in Information and Computing Science (Mathematics)
*GPA :*3.65/4.30 *Average Score:*86.71
Selected Courses : Mathematical Analysis (95/100) Probability Theory (93/100) Operations Research (85/100) Mathematical Modeling (A+) Analysis and Practice of the Data (86/100) Introduction to Deep Learning (90/100)

Experience

Theoretical Analysis of Generalization of Diffusion Models — Jul. 2023 – Present

Summer Research Internship
Mentor: Difan Zou University of Hong Kong

Brief introduction:
The current theoretical analysis of the generalization capability of diffusion models is significantly lacking. Thus this project aims to dive deep into the core of diffusion models' ability to generate novel images as well as their generalization.

- Explored the capability of DDPM trained on tiny dataset to generate novel samples not in the training set.
- Analyzed the impact of varying network capacities on generalization.
- Compared the trained noise ε_θ and the existing empirical optima of training objective in the literature.

Privacy Preserving Diffusion Model — Feb. 2023 – Present

Undergraduate Student Research Program
Mentor: Jingrun Chen University of Science and Technology of China

Brief introduction:
This program focuses on two critical challenges faced by diffusion models: slow sampling speed and their inclination to replicate training data. The former impedes its practical application in real-world scenarios, while the latter may potentially trigger copyright disputes and privacy leakage.

- Reimplemented several generative models like DDPM, DDIM, and DPM-Solver.
- Investigated fast sampling methods, including ode solver and progressive distillation.
- Explored diffusion models' memorization of training data based on ViT and tried to mitigate it during training process.

PISA Data Analysis — Mar. 2022 – May. 2022

Data Analysis Course Project University of Science and Technology of China

Brief introduction: Based on the dataset released by PISA (Programme for International Student Assessment) in 2015, pre-processing the data, extracting features, predicting student repetition, and students' mathematical ability.

- Pre-processed high dimensional data.
- Implemented Naive Bayes classifier from scratch (accuracy over 98% on validation set).
- Predicted 'mathematical ability' (an unknown feature) via different machine learning methods such as Linear Regression, Random Forest, Naive Bayes and KNN.

Skills

- Programming Languages: Python(Pytorch), C++, MATLAB, Mathematica
- English: TOEFL – 92(R29,L19,S20,W24), CET4 – 645, CET6 – 574, EF SET – C2
- Platform: Windows

Honors and Awards

Excellent Student Scholarship — 2021, 2023
3rd Prize, China Undergraduate Mathematical Contest in Modeling (CUMCM) — 2021
3rd Prize (CMC), Chinese Mathematics Competition — 2022